



# The Changing Face of Genomics

## Citation

Kellis, Manolis. 2004. The changing face of genomics. *Genome Biology* 5(5): 324.

## Published Version

<http://genomebiology.com/2004/5/5/324>

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4553780>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

# Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# The changing face of genomics

Manolis Kellis

Address: The Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02138, and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: manoli@mit.edu

Published: 30 April 2004

*Genome Biology* 2004, **5**:324

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/5/324>

© 2004 BioMed Central Ltd

---

A report on the 5th annual Advances in Genome Biology and Technology (AGBT) and Automation in DNA Mapping and Sequencing (AMS) meeting, Marco Island, USA, 4-7 February 2004.

---

The annual meeting on Advances in Genome Biology and Technology was very different this year - in contrast to previous years, only a handful of talks covered the latest large-scale sequencing projects and the next species to be sequenced. This meeting took for granted that we can sequence, assemble and align complete genomes - achievements that only a few years ago seemed daunting, if not unthinkable. The focus of the meeting has instead shifted towards the new challenges in genomics, particularly in the areas of gene regulation, cell dynamics and genome evolution.

## Cell regulation and organism development

Given the primary sequence of a species, a major goal of current genomics efforts is to understand the regulatory mechanisms and control circuitry of the cell. Towards this goal, Rick Young (Massachusetts Institute of Technology and Whitehead Institute, Cambridge, USA) presented the completion of the yeast protein-DNA interaction map. Using chromatin immuno-precipitation (ChIP) technology in combination with microarray 'chips' containing all intergenic regions, his group has undertaken a genome-wide study of the targets of all the roughly 200 transcriptional regulators in yeast under multiple environmental conditions. They characterized the sequence-specificity of these regulators using numerous motif-discovery tools, together with evolutionary sequence conservation and protein structure information. The resulting regulatory map revealed general principles of regulation in yeast, including the organizational architectures of promoter regions (single motif, multiple sites, multiple regulators or factor combinations), and the

different types of regulatory response to environmental changes (off/on, invariant, expanded or altered). The Young group also studied 20 chromatin regulators that do not directly recognize DNA sequences, but instead rely on their association with transcription-factor partners for binding and can keep a record of transcription by maintaining chromatin state. The Young lab is now moving this technology into studying transcriptional-regulator binding in the human genome, which will have applications to understanding diseases from diabetes to cancer.

Michael Levine (University of California, Berkeley, USA) presented work aiming to understand the *cis*-regulatory circuitry of promoter regions in the fruitfly. He proposed that enhancer complexity might be a better measure of organismal complexity than overall gene count, and noted that the majority of enhancer elements act cooperatively in higher eukaryotes - autonomously-acting elements would in fact be the exceptions. Levine's group has studied the promoter architecture of *Drosophila* developmental genes that respond to different levels of the concentration gradient of the developmental protein Dorsal. Five activation levels, created by the combinatorial action of a three-response-level activator and a repressor, were detected for this set of genes. Levine and colleagues searched for conserved sequence elements in the promoter regions of genes belonging to the same activation level and discovered a common 'grammar' in the organization of three basic enhancer elements. Searching for a similar regulatory grammar in the mosquito genome revealed ten genes with similar putative regulatory clusters, two of which contain the same architecture, despite 230 million years of divergence.

Both Young and Levine have benefited from earlier work - classical studies of gene function in yeast and known developmental genes in the fly - that have set the foundations for future studies. In this spirit, Nancy Hopkins (Massachusetts Institute of Technology, Cambridge, USA) presented her research program that aims to systematically

identify all developmental genes in the zebrafish. Zebrafish is an appealing model for studying early vertebrate development because of its transparent body and the short time - a mere four days - between fertilization and free-swimming larval stages. As reaching the adult stage takes another four months, following multiple generations can, however, be prohibitively slow. To face this challenge, Hopkins and colleagues have used insertional mutagenesis to create mosaic parents whose germ cells each contain a different mutation. Following the fish with developmental defects and classifying each mutation has allowed her group to screen 32,000 founder fish, identify 550 mutants and 390 loci, 298 of which have human homologs. As many as 20% of these genes have no previously known biochemical function, providing a great starting point for experimentation and new biological discoveries. Additionally, the systematic approach allows one to estimate the total number of developmental genes, which Hopkins sets at 1,600, of which 25% have already been isolated. The cost of identifying additional genes increases as the study approaches saturation, however, and her group is not planning to pursue the systematic discovery phase of the work. They are currently working on understanding the genes identified and have revealed important new insights about genes involved in kidney, jaw, liver and myeloid cell formation.

### Protein interactions and network evolution

Beyond the identification of genes involved in a developmental process lies the major challenge of understanding the genes' dynamic patterns of behavior during development. Josh LaBaer (Harvard Medical School, Boston, USA) presented his lab's proteomics work enabling such a pursuit, which they have approached by developing a 'protein-expression clone repository' that contains full-length protein-coding sequences for every gene in a number of model organisms, including yeast, bacteria and human. These protein-coding sequences are inserted into 'master' clone vectors and can be easily transferred to specialized vectors for expression studies, tagging with green fluorescent protein (GFP) to study localization, two-hybrid assays to determine protein interactions, or expression in constructs to detect protein modification states by mass spectroscopy. The system architecture is designed to be flexible, modular, reliable, comprehensive and catalogued. Mark Vidal (Harvard Medical School, Boston, USA) described a similar system for understanding the worm proteome. His lab is now mapping the localization of all 19,000 worm gene products across development, and so building 'chronograms' of gene expression from the head to the tail. The patterns of protein localization can then be clustered across space and time, thus constructing the dynamic aspect of the protein-interaction network of the worm. Vidal's group has currently completed 10% of the interactome matrix, and is moving towards completeness with the goal of understanding not only the components of the interactome but also protein behaviors during complex organismal tasks.

The protein-interaction network of a species provides the foundation for understanding the organism's responses to environmental changes and developmental signals. Across evolutionary time, these responses change and the regulatory circuits shift towards new ones. Lisa Stubbs (Lawrence Livermore National Laboratory, Livermore, USA) presented work aimed at understanding the evolution of such regulatory mechanisms across species. Her group studied the KRAB-ZNF family of chromatin-interacting zinc-finger transcriptional regulators. These proteins arose 400 million years ago, after divergence of land vertebrates from fish, and have recently undergone lineage-specific expansions in the human and mouse lineages via tandem duplications and deletions. Researchers shy away from using traditional comparative genomics tools for the analysis of such expanding gene families, as orthologous pairs are hard to determine. In the absence of orthology information, Stubbs constructed multiple alignments of all paralogous gene copies within each species in turn, rather than across species. This clever methodology, although atypical, yielded biologically meaningful intergenic sequence elements that are highly conserved across paralogs, and which can be shown, on the basis of reporter assays, to indeed act as enhancers. This approach also allowed the identification of species-specific elements that have arisen since the divergence of the mouse and human lineages. Stubbs' group is now working out the protein-level differences between paralogs, with the goal of understanding how structural changes affect function, in particular with respect to tissue-specific regulation and parental imprinting.

### Human divergence and diversity

Differences between more closely related species can be much more subtle than the protein family expansions observed between human and mouse. For example, divergence between human and chimp shows as few as 12 changes every 1,000 nucleotides, which makes biological signal discovery a real challenge. Mike Zody (Massachusetts Institute of Technology and Broad Institute, Cambridge, USA) presented a very interesting way to use such a close relative to reveal insights into recent evolutionary history. Assuming that neutral divergence between species and neutral diversity within species are driven by the same underlying mutational mechanisms, Zody and colleagues used human-chimp divergence information to model the background mutation rate for each region of the genome. Using this information, they were able to distinguish regions of low within-human diversity that were due to selective sweeps (where a favorable mutation becomes incorporated into the genome very quickly), rather than simply to a lower mutation rate. These regions of low diversity were confirmed by genotyping nucleotide polymorphisms in humans of different ethnicity and building allele-frequency profiles. *FOXP2*, a well-known gene involved in language development, showed only a moderate signal for selection, whereas some regions that are devoid of annotated genes showed very strong evidence of

selection, raising new questions about what type of genetic elements may be under selection in the human genome.

With the aim of understanding human variation and how genetic differences may relate to phenotypic differences, Joanna Mountain (Stanford University, USA) presented her work relating language similarities to population polymorphisms. In particular, she has studied the genetic relationships of two geographically isolated populations in Africa, both of which speak with clicking languages (Khoisan). Using a multitude of metrics across various loci and populations, she found that the genetic data clearly support an evolutionarily distinct group for each of the two languages. To explain the results, Mountain put forward three possibilities: that the two languages arose independently, that they traveled across populations, or that instead the ancestral language was click-based and all intervening groups lost the click sounds. One advantage to such click languages might be the ability to communicate while hunting without alerting the prey. The use of modern genomics tools in the population genetics of language evolution illustrates how diverse the genomics field has become, and that, in fact, the maturity of genomics as a science has moved its applications well beyond the boundaries of the modern laboratory.

Overall, the meeting showcased a wide range of innovative talks, combining ground-breaking technological inventions with important scientific applications. The attendance was unusually low this year, especially on the industry side, witnessing the tight economic situation in the USA and internationally. At the same time, the changing focus of the meeting is evidence of a maturing field. Genomics has mastered its initial challenges, and is now extending its arms to embrace a growing number of fundamental questions in the study of life.